Introduction to Data Assimilation

or alternatively

Introduction to Estimation Theory

Ricardo Todling

Global Modeling and Assimilation Office, NASA/GSFC

Goddard 2003 Summer School

http://gmao.gsfc.nasa.gov/DAO_people/todling/todling.html

Outline

1. Objectives

2. Concepts of probabilistic estimation

3. Example: Estimation of a constant vector

4. The probabilistic approach to filtering

5. The probabilistic approach to smoothing

6. Illustrations

Illustration 1: Data Assimilation for Chaotic Dynamics

Dynamical System: Lorenz (1963)

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= \rho x - y - xz \\ \dot{z} &= xy - \beta z \end{aligned}$$

Chaotic for the following parameters:

$$\sigma = 10 \quad \rho = 28 \quad \beta = 8/3$$

Unstable equilibrium points:

$$(0,0,0)$$

$$(\pm\sqrt{\beta(\rho - 1)}, \pm\sqrt{\beta(\rho - 1)}, -1)$$

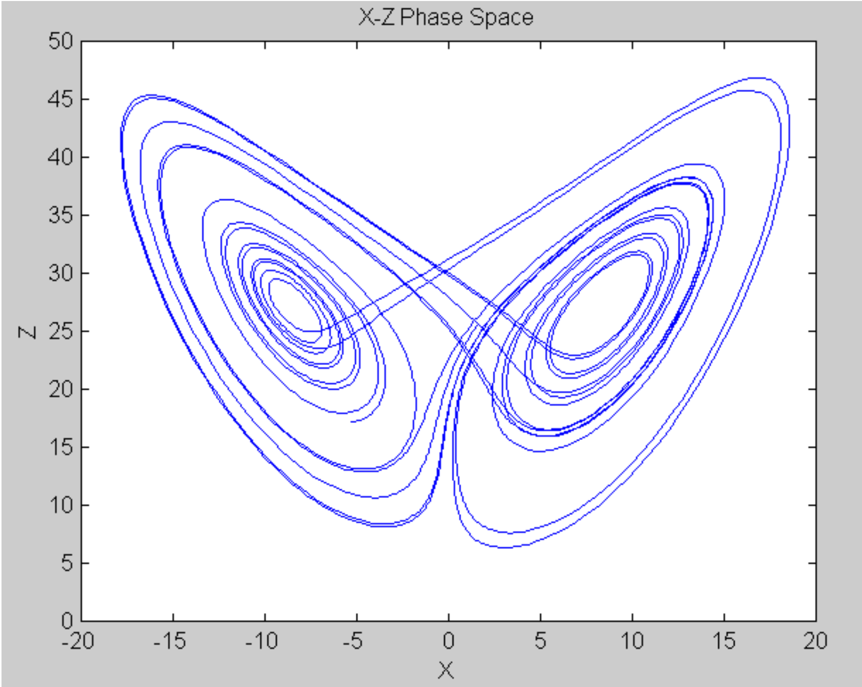
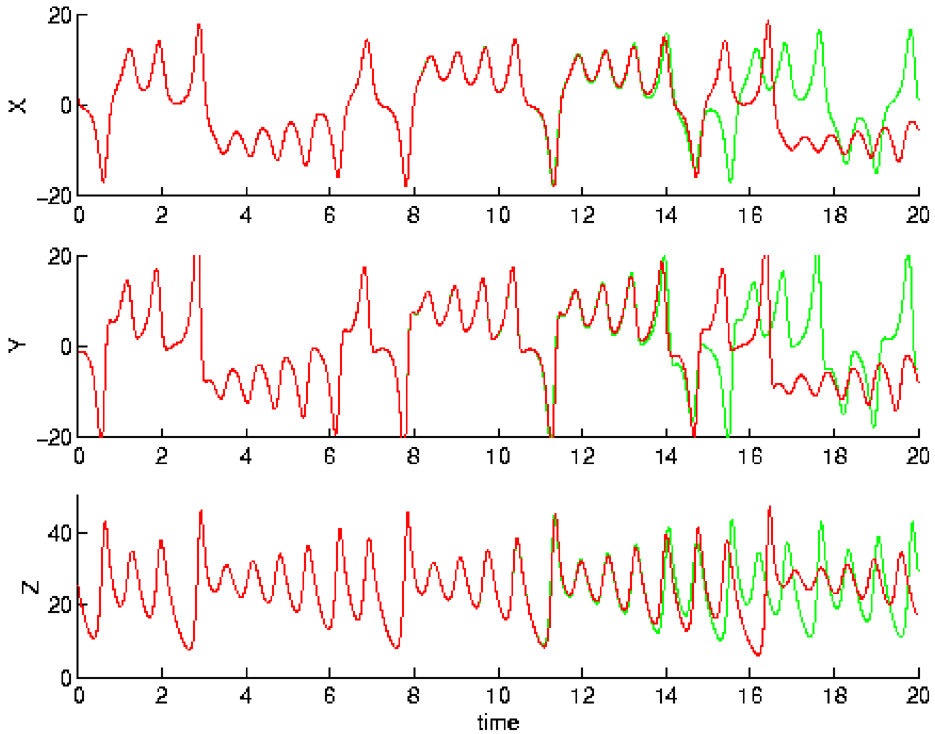


Illustration 1(cont.): Data Assimilation for Chaotic Dynamics

What does a tiny initial perturbation do to prediction?

$\sigma(0) = 10^{-6}$

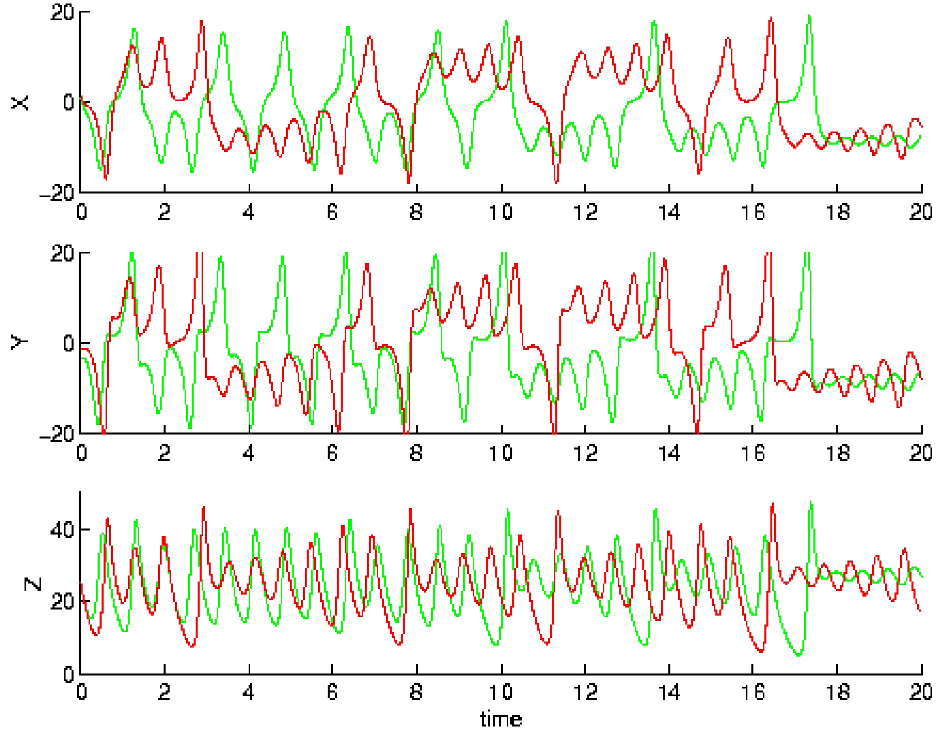


Answer: Cause some (chaotic) trouble!

Illustration 1(cont.): Data Assimilation for Chaotic Dynamics

What about a not-so-tiny initial perturbation?

$\sigma(0) = 1$



<u>Answer</u>: It causes a lot of trouble! The two runs started from initial conditions differing by about one percent in magnitude. You can think of the red line as being the true state evolution and the green line as being the predicted state. In this case, the prediction becomes useless very quickly. The solution to this problem is to assimilate observations.

1. *Objectives*

The main objective of this lecture is to present a summary of some of the methods most commonly used for state estimation.

What I hope you take from this lecture is that:

- ▷ The *probabilistic approach* allows for the proper description of most (if not all) methods.

- ▷ In practice, most methods used in atmospheric and oceanic data assimilation boil down to slightly different versions of *least-squares*.

- ▷ In the end, what really matters are things like:
  - off-line and on-line quality control
  - removal of both model and observation biases
  - proper usage of observations, that is, they should be used at right time, be given proper representativeness error characteristics
  - watch for details ...

- ▷ Always remember ... *adaptive procedures are robust*.

2.1 Bayesian Approach

Define a cost (risk) function:

$$
\begin{aligned}
\mathcal{J}(\hat{\mathbf{w}}) &\equiv \mathcal{E}\{J(\tilde{\mathbf{w}})\} \\
&= \int_{-\infty}^{\infty} J(\tilde{\mathbf{w}})\, p_{\mathbf{w}}(\mathbf{w})\, d\mathbf{w} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J(\tilde{\mathbf{w}})\, p_{\mathbf{wz}}(\mathbf{w},\mathbf{z})\, d\mathbf{z}\, d\mathbf{w}
\end{aligned}
$$

where

$\mathbf{w}$	true state vector
$\mathbf{z}$	observation vector
$\hat{\mathbf{w}}$	state estimate vector
$\tilde{\mathbf{w}}$	error estimate vector: $\hat{\mathbf{w}} - \mathbf{w}$
$J(\tilde{\mathbf{w}})$	measure of accuracy
$p_{\mathbf{w}}(\mathbf{w})$	marginal pdf of $\mathbf{w}$
$p_{\mathbf{wz}}(\mathbf{w},\mathbf{z})$	joint pdf between $\mathbf{w}$ and $\mathbf{z}$

Note: Not all function J's are satisfactory cost functions.

Tip 1: Conditional Probability

$$p_{\mathbf{wz}}(\mathbf{w}|\mathbf{z}) \equiv \frac{p_{\mathbf{wz}}(\mathbf{w},\mathbf{z})}{p_{\mathbf{z}}(\mathbf{z})}$$

Tip 2: Bayes rule

$$p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z}) = \frac{p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})p_{\mathbf{w}}(\mathbf{w})}{p_{\mathbf{z}}(\mathbf{z})}$$

Two Examples of Cost Functions

(a) The quadratic cost:

$$J = ||\tilde{\mathbf{w}}||_{\mathbf{E}}^2 = \tilde{\mathbf{w}}^T \mathbf{E} \tilde{\mathbf{w}}$$

(b) The uniform cost:

$$J = \begin{cases} 0, & ||\tilde{\mathbf{w}}|| < \epsilon \\ 1/2\epsilon, & ||\tilde{\mathbf{w}}|| \geq \epsilon \end{cases}$$

A desirable property of an estimate is that it be unconditionally unbiased, that is,

$$\mathcal{E}\{\hat{\mathbf{w}}\} = \mathcal{E}\{\mathbf{w}\}$$

Sometimes we could have an estimate conditionally unbiased:

$$\mathcal{E}\{\hat{\mathbf{w}}|\mathbf{w}\} = \mathbf{w}$$

2.2 Minimum Variance Estimation

In this case we use the quadratic cost function to get:

$$\mathcal{J}_{\mathrm{MV}}(\hat{\mathbf{w}}) = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} (\mathbf{w} - \hat{\mathbf{w}})^T \mathbf{E} (\mathbf{w} - \hat{\mathbf{w}})\, p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \right\} p_{\mathbf{z}}(\mathbf{z})\, d\mathbf{z}$$

Or, identifying the kernel as the conditional Bayes cost:

$$\mathcal{J}_{\mathrm{MV}}(\hat{\mathbf{w}}|\mathbf{z}) \equiv \int_{-\infty}^{\infty} (\mathbf{w} - \hat{\mathbf{w}})^T \mathbf{E} (\mathbf{w} - \hat{\mathbf{w}})\, p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w}$$

Minimization of the cost $\mathcal{J}_{\mathrm{MV}}(\hat{\mathbf{w}}|\mathbf{z})$ gives

$$\begin{aligned} \mathbf{0} &= \left. \frac{\partial \mathcal{J}_{\mathrm{MV}}(\hat{\mathbf{w}}|\mathbf{z})}{\partial \hat{\mathbf{w}}} \right|_{\hat{\mathbf{w}}=\hat{\mathbf{w}}_{\mathrm{MV}}} \\ &= \left. -\, 2\, \mathbf{E} \int_{-\infty}^{\infty} (\mathbf{w} - \hat{\mathbf{w}})\, p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \right|_{\hat{\mathbf{w}}=\hat{\mathbf{w}}_{\mathrm{MV}}} \end{aligned}$$

And noticing that p is a pdf, it follows that

$$\begin{aligned} \hat{\mathbf{w}}_{\mathrm{MV}}(\mathbf{z}) &= \int_{-\infty}^{\infty} \mathbf{w} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \\ &= \mathcal{E}\{\mathbf{w}|\mathbf{z}\} \end{aligned}$$

Conclusion: the estimate with minimum variance is the conditional mean.

- this estimate is unbiased
- this estimate is indeed the minimum of the cost function

2.3 *Maximum a posteriori Probability Estimation*

Using now the uniform cost function we have

$$\mathcal{J}_U(\hat{\mathbf{w}}) = \int_{-\infty}^{\infty} \frac{1}{2\epsilon}\left\{1 - \int_{\hat{\mathbf{w}}-\epsilon}^{\hat{\mathbf{w}}+\epsilon} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w}\right\} p_{\mathbf{z}}(\mathbf{z})\, d\mathbf{z}$$

To minimize $\mathcal{J}_U$ with respect to $\hat{\mathbf{w}}$, the first term gives no relevant contribution, thus

$$\mathcal{J}_U(\hat{\mathbf{w}}) \sim -(1/2\epsilon)\int_{-\infty}^{\infty}\left\{\int_{\hat{\mathbf{w}}-\epsilon}^{\hat{\mathbf{w}}+\epsilon} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w}\right\} p_{\mathbf{z}}(\mathbf{z})\, d\mathbf{z}\,.$$

or yet, we can minimize the conditional Bayes cost

$$\mathcal{J}_U(\hat{\mathbf{w}}|\mathbf{z}) \equiv -(1/2\epsilon)\int_{\hat{\mathbf{w}}-\epsilon}^{\hat{\mathbf{w}}+\epsilon} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w}$$

As $\epsilon \to 0$, the mean value theorem gives

$$\mathcal{J}_U(\hat{\mathbf{w}}|\mathbf{z}) = -\,p_{\mathbf{w}|\mathbf{z}}(\hat{\mathbf{w}}|\mathbf{z})$$

Conclusion: The maximum a posteriori estimate is obtained by maximizing the conditional pdf, that is,

$$\left.\frac{\partial \ln[p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})p_{\mathbf{w}}(\mathbf{w})]}{\partial \mathbf{w}}\right|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{MAP}}} = 0$$

or yet

$$\left.\frac{\partial p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})p_{\mathbf{w}}(\mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{MAP}}} = 0$$

- this estimate is NOT guaranteed to be unbiased

2.4 Maximum Likelihood Estimation

In ML estimation we assume the *a priori* information is unknown. Suppose for the moment that the *a priori* pdf is $\mathcal{N}(\boldsymbol{\mu}_{\mathbf{w}}, \mathbf{P}_{\mathbf{w}})$, then

$$\ln p_{\mathbf{w}}(\mathbf{w}) = -\ln[(2\pi)^{n/2}|\mathbf{P}_{\mathbf{w}}|^{1/2}] - \frac{1}{2}\left[(\mathbf{w}-\boldsymbol{\mu}_{\mathbf{w}})^T\mathbf{P}_{\mathbf{w}}^{-1}(\mathbf{w}-\boldsymbol{\mu}_{\mathbf{w}})\right]$$

Hence,

$$\frac{\partial \ln p_{\mathbf{w}}(\mathbf{w})}{\partial \mathbf{w}} = -\mathbf{P}_{\mathbf{w}}^{-1}(\mathbf{w}-\boldsymbol{\mu}_{\mathbf{w}})$$

indicating that lack of information implies infinite variance, $\mathbf{P}_{\mathbf{w}} \to \infty$, or yet $\mathbf{P}_{\mathbf{w}}^{-1} \to \mathbf{0}$. Consequently, the maximum likelihood estimate of $\mathbf{w}$ can be obtained by

$$\begin{aligned} \mathbf{0} &= \left[\frac{\partial \ln p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})}{\partial \mathbf{w}} + \frac{\partial \ln p_{\mathbf{w}}(\mathbf{w})}{\partial \mathbf{w}}\right]\Bigg|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{MAP}}} \\ &= \frac{\partial \ln p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})}{\partial \mathbf{w}}\Bigg|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{ML}}} \end{aligned}$$

or equivalently,

$$\frac{\partial p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})}{\partial \mathbf{w}}\Bigg|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{ML}}} = \mathbf{0}$$

- $\hat{\mathbf{w}}_{\mathrm{ML}}$ can be referred to as the most likely estimate
- This estimate is NOT guaranteed to be unbiased.
- The estimate obtained this way is NOT Bayesian.

Quick Recap

Minimum variance estimate:

$$\begin{aligned}\hat{\mathbf{w}}_{\mathrm{MV}}(\mathbf{z}) &= \int_{-\infty}^{\infty} \mathbf{w}\, p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} \\ &= \mathcal{E}\{\mathbf{w}|\mathbf{z}\}\end{aligned}$$

Maximum *a posteriori* probability estimate:

$$\left.\frac{\partial p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})\, p_{\mathbf{w}}(\mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{MAP}}} = \mathbf{0}$$

Maximum likelihood estimate (max *a priori* pdf):

$$\left.\frac{\partial p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{ML}}} = \mathbf{0}$$

3. *Example: Estimation of a Constant Vector*

Consider the observational process

$$\mathbf{z} = \mathbf{H}\mathbf{w} + \mathbf{b}^o$$

where $\mathbf{w}$ is an n–vector, $\mathbf{z}$ and $\mathbf{b}^o$ are m–vectors, and $\mathbf{H}$ is an $m \times n$ matrix.

Assumptions: $\mathbf{w}$ and $\mathbf{b}^o$ are independent and Gaussian distributed, that is, $\mathbf{w} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{P})$, and $\mathbf{b}^o \sim \mathcal{N}(0, \mathbf{R})$.

Problem: What do the three estimates studied previously correspond to in this case?

For the MV estimate we need to determined the *a posteriori* pdf $p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})$ (Bayes rule):

$$p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z}) = \frac{p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})p_{\mathbf{w}}(\mathbf{w})}{p_{\mathbf{z}}(\mathbf{z})}$$

consequently we need to determine each one of the pdf's above.

Linear transformations of Gaussian distributed variables result in Gaussian distributed variables. Therefore,

$$p_{\mathbf{z}}(\mathbf{z}) = \frac{1}{(2\pi)^{m/2}|\mathbf{P}_{\mathbf{z}}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})^T \mathbf{P}_{\mathbf{z}}^{-1}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})\right]$$

where $\boldsymbol{\mu}_{\mathbf{z}}$ and $\mathbf{P}_{\mathbf{z}}$ correspond to the mean and covariance of the random variable $\mathbf{z}$, respectively.

Applying the ensemble average operator and using the definition of covariance:

$$\boldsymbol{\mu}_{\mathbf{z}} = \mathcal{E}\{\mathbf{H}\mathbf{w}\} + \mathcal{E}\{\mathbf{b}^o\} = \mathbf{H}\boldsymbol{\mu}$$

and also,

$$\begin{aligned}
\mathbf{P}_{\mathbf{z}} &= \mathcal{E}\{(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})^T\} \\
&= \mathcal{E}\{[(\mathbf{H}\mathbf{w}+\mathbf{b}^o)-\mathbf{H}\boldsymbol{\mu})][(\mathbf{H}\mathbf{w}+\mathbf{b}^o)-\mathbf{H}\boldsymbol{\mu})]^T\} \\
&= \mathcal{E}\{[(\mathbf{H}\mathbf{w}-\mathbf{H}\boldsymbol{\mu})-\mathbf{b}^o][(\mathbf{H}\mathbf{w}-\mathbf{H}\boldsymbol{\mu})-\mathbf{b}^o]^T\} \\
&= \mathbf{H}\mathcal{E}\{(\mathbf{w}-\boldsymbol{\mu})(\mathbf{w}-\boldsymbol{\mu})^T\}\mathbf{H}^T + \mathcal{E}\{\mathbf{b}^o\mathbf{b}^{oT}\} \\
&\quad + \mathbf{H}\mathcal{E}\{(\mathbf{w}-\boldsymbol{\mu})\mathbf{b}^{oT}\} + \mathcal{E}\{\mathbf{b}^o(\mathbf{w}-\boldsymbol{\mu})^T\}\mathbf{H}^T .
\end{aligned}$$

Noticing that $\mathbf{w}$ and $\mathbf{b}^o$ are independent $\mathcal{E}\{\mathbf{w}\mathbf{b}^{oT}\} = \mathbf{0}$, and that $\mathbf{b}^o$ has zero mean, it follows that

$$\mathbf{P}_{\mathbf{z}} = \mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}$$

Consequently,

$$\begin{aligned}
p_{\mathbf{z}}(\mathbf{z}) &= \frac{1}{(2\pi)^{m/2}|(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})|^{1/2}} \\
&\quad \times \exp\left[-\frac{1}{2}(\mathbf{z}-\mathbf{H}\boldsymbol{\mu})^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{z}-\mathbf{H}\boldsymbol{\mu})\right]
\end{aligned}$$

It remains to determine the conditional pdf $p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})$. This distribution is also Gaussian, and can be written as

$$p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w}) = \frac{1}{(2\pi)^{m/2}|\mathbf{P}_{\mathbf{z}|\mathbf{w}}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}|\mathbf{w}})^T \mathbf{P}_{\mathbf{z}|\mathbf{w}}^{-1}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}|\mathbf{w}})\right]$$

Analogously to what we have just done above,

$$\boldsymbol{\mu}_{\mathbf{z}|\mathbf{w}} = \mathcal{E}\{\mathbf{H}\mathbf{w}|\mathbf{w}\} + \mathcal{E}\{\mathbf{b}^o|\mathbf{w}\} = \mathbf{H}\mathbf{w}$$

and

$$\begin{aligned}
\mathbf{P}_{\mathbf{z}|\mathbf{w}} &= \mathcal{E}\{(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}|\mathbf{w}})(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}|\mathbf{w}})^T|\mathbf{w}\} \\
&= \mathcal{E}\{\,[(\mathbf{H}\mathbf{w}+\mathbf{b}^o)-\mathbf{H}\mathbf{w})]\,[(\mathbf{H}\mathbf{w}+\mathbf{b}^o)-\mathbf{H}\mathbf{w})]^T\,|\mathbf{w}\} \\
&= \mathcal{E}\{\mathbf{b}^o\mathbf{b}^{oT}|\mathbf{w}\} \\
&= \mathcal{E}\{\mathbf{b}^o\mathbf{b}^{oT}\} \\
&= \mathbf{R}\,.
\end{aligned}$$

Therefore,

$$p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w}) = \frac{1}{(2\pi)^{m/2}|\mathbf{R}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{z}-\mathbf{H}\mathbf{w})^T\mathbf{R}^{-1}(\mathbf{z}-\mathbf{H}\mathbf{w})\right]$$

which is the conditional probability of $\mathbf{z}$ given $\mathbf{w}$.

Combining the previous results in Bayes rule for pdf's:

$$p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z}) = \frac{|\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}|^{1/2}}{(2\pi)^{n/2}|\mathbf{P}|^{1/2}|\mathbf{R}|^{1/2}} \exp[-\frac{1}{2}J]$$

where J is defined as,

$$\begin{aligned} J(\mathbf{w}) \equiv\ & (\mathbf{z} - \mathbf{H}\mathbf{w})^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{w}) + (\mathbf{w} - \boldsymbol{\mu})^T\mathbf{P}^{-1}(\mathbf{w} - \boldsymbol{\mu}) \\ & - (\mathbf{z} - \mathbf{H}\boldsymbol{\mu})^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{z} - \mathbf{H}\boldsymbol{\mu}) \end{aligned}$$

This quantity J can also be written in the following more compact form:

$$J(\mathbf{w}) = (\mathbf{w} - \hat{\mathbf{w}})^T\mathbf{P}_{\tilde{\mathbf{w}}}^{-1}(\mathbf{w} - \hat{\mathbf{w}})$$

where $\mathbf{P}_{\tilde{\mathbf{w}}}^{-1}$ is given by

$$\mathbf{P}_{\tilde{\mathbf{w}}}^{-1} = \mathbf{P}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H},$$

the vector $\hat{\mathbf{w}}$ is given by

$$\hat{\mathbf{w}} = \mathbf{P}_{\tilde{\mathbf{w}}}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z} + \mathbf{P}^{-1}\boldsymbol{\mu})$$

and the reason for using the subscript $\tilde{\mathbf{w}}$ for the matrix $\mathbf{P}_{\tilde{\mathbf{w}}}$, indicating a relationship with the estimation error, will soon become clear.

We are now ready to derived the desired estimates.

The minimum variance estimate is given by the conditional mean of the *a posteriori* pdf, that is,

$$\hat{\mathbf{w}}_{\mathrm{MV}} = \int_{-\infty}^{\infty} \mathbf{w} p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})\, d\mathbf{w} = \hat{\mathbf{w}}$$

The maximum *a posteriori* probability estimate is the one that maximizes $p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})$, and is easily identified to be

$$\hat{\mathbf{w}}_{\mathrm{MAP}} = \hat{\mathbf{w}} = (\mathbf{P}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z} + \mathbf{P}^{-1}\boldsymbol{\mu})$$

The maximum likelihood estimate can be determined by maximizing the pdf $p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})$, that is,

$$\mathbf{0} = \left.\frac{\partial p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{w}=\hat{\mathbf{w}}_{\mathrm{ML}}} = \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\hat{\mathbf{w}}_{\mathrm{ML}})$$

that is,

$$\hat{\mathbf{w}}_{\mathrm{ML}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z}$$

which is, in principle, distinct from the estimates obtained above.

The MV and MAP estimates can be reduced to the ML estimate by taking $\mathbf{P}^{-1} = \mathbf{0}$, that is, when no statistical information on $\mathbf{w}$ is available:

$$\hat{\mathbf{w}}_{\mathrm{MV}}|_{\mathbf{P}^{-1}=\mathbf{0}} = \hat{\mathbf{w}}_{\mathrm{MAP}}|_{\mathbf{P}^{-1}=\mathbf{0}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z} = \hat{\mathbf{w}}_{\mathrm{ML}}$$

Quick Recap

Observations: $\mathbf{z} = \mathbf{H}\mathbf{w} + \mathbf{v}$

Want to determine: $p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z})$

when $\mathbf{w} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{P})$, and $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, we find:

$$p_{\mathbf{w}|\mathbf{z}}(\mathbf{w}|\mathbf{z}) \,\alpha\, \exp[-\frac{1}{2}(\mathbf{w} - \hat{\mathbf{w}})^T \mathbf{P}_{\tilde{\mathbf{w}}}^{-1}(\mathbf{w} - \hat{\mathbf{w}})]$$

where

$$\mathbf{P}_{\tilde{\mathbf{w}}}^{-1} = \mathbf{P}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H},$$

and

$$\hat{\mathbf{w}} = \mathbf{P}_{\tilde{\mathbf{w}}}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z} + \mathbf{P}^{-1}\boldsymbol{\mu})$$

Estimation Results:

$$\hat{\mathbf{w}}_{\mathrm{MV}} = \hat{\mathbf{w}}_{\mathrm{MAP}} = \hat{\mathbf{w}}$$

$$\hat{\mathbf{w}}_{\mathrm{ML}} = \hat{\mathbf{w}}_{\mathrm{MV}}\big|_{\mathbf{P}^{-1}=0} = \hat{\mathbf{w}}_{\mathrm{MAP}}\big|_{\mathbf{P}^{-1}=0}$$

Remarks

- All estimates above result in a *linear combination* of the observations.

- The MAP estimate can be obtained by minimizing the alternative cost function

  $$J_{\mathrm{MAP}}(\mathbf{w}) \equiv (\mathbf{z}-\mathbf{H}\mathbf{w})^T\mathbf{R}^{-1}(\mathbf{z}-\mathbf{H}\mathbf{w}) + (\mathbf{w}-\boldsymbol{\mu})^T\mathbf{P}^{-1}(\mathbf{w}-\boldsymbol{\mu}),$$

  which amounts to noticing that the pdf $p_{\mathbf{z}}(\mathbf{z})$ does not play any role in the maximization of the *a posteriori* pdf.

- Similarly, the ML estimate can be obtained by minimizing the following cost function:

  $$J_{\mathrm{ML}}(\mathbf{w}) \equiv (\mathbf{z}-\mathbf{H}\mathbf{w})^T\mathbf{R}^{-1}(\mathbf{z}-\mathbf{H}\mathbf{w}),$$

  and corresponding estimate is biased.

- In general there is no guarantee these three estimates coincide. In the case just considered they only coincide after knowledge on the prior is ignored in the MV and MAP results.

The Least-Squares (LS) Connection

Case I: No prior information on $\mathbf{w}$ is available.

Minimization of the cost function

$$J_{\mathrm{LS}}(\hat{\mathbf{w}}) = (\mathbf{z} - \mathbf{H}\hat{\mathbf{w}})^T \tilde{\mathbf{R}}^{-1} (\mathbf{z} - \mathbf{H}\hat{\mathbf{w}})$$

results in

$$\hat{\mathbf{w}}_{\mathrm{LS}} = (\mathbf{H}^T \tilde{\mathbf{R}}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \tilde{\mathbf{R}}^{-1} \mathbf{z}$$

which is identical to the ML (MV/MAP) estimate(s) if $\tilde{\mathbf{R}} = \mathbf{R}$. In general, however, the LS solution can be shown to always be less accurate than that of ML (MV/MAP).

Case II: Some information on $\mathbf{w}$ is available.

The cost function to be minimized is now

$$J_{\mathrm{LSP}}(\hat{\mathbf{w}}) = (\boldsymbol{\mu} - \hat{\mathbf{w}})^T \tilde{\mathbf{P}}^{-1} (\boldsymbol{\mu} - \hat{\mathbf{w}}) + (\mathbf{z} - \mathbf{H}\hat{\mathbf{w}})^T \tilde{\mathbf{R}}^{-1} (\mathbf{z} - \mathbf{H}\hat{\mathbf{w}})$$

with minimum achieved for

$$\hat{\mathbf{w}}_{\mathrm{LSP}} = (\tilde{\mathbf{P}}^{-1} + \mathbf{H}^T \tilde{\mathbf{R}}^{-1} \mathbf{H})^{-1} (\mathbf{H}^T \tilde{\mathbf{R}}^{-1} \mathbf{z} + \tilde{\mathbf{P}}^{-1} \boldsymbol{\mu})$$

which is identical to the MV/MAP estimate if $\tilde{\mathbf{R}} = \mathbf{R}$ and $\tilde{\mathbf{P}} = \mathbf{P}$. In general, however, the LSP solution can be shown to be always less accurate than that of MV/MAP.

4. *The Probabilistic Approach to Filtering*

Let us indicate by $\mathbf{W}^o_k = \{\mathbf{w}^o_1, \cdots, \mathbf{w}^o_{k-1}, \mathbf{w}^o_k\}$, the set of all observations up to and including time t_k. Similarly, let us indicate by $\mathbf{W}^t_k = \{\mathbf{w}^t_1, \cdots, \mathbf{w}^t_{k-1}, \mathbf{w}^t_k\}$ the set of all true states of the underlying system up to time t_k.

Knowledge of the pdf of the true state over the entire time period given all observations over the same period would allow us to calculate an estimate of the trajectory of the system over the time period. Therefore, calculation of the following pdf

$$p(\mathbf{W}^t_k|\mathbf{W}^o_k)$$

is desirable. But, before seeking a system trajectory estimate, let us seek an estimate of the state of the system only at time t_k. For that, the relevant pdf is

$$\begin{aligned}
p(\mathbf{w}^t_k|\mathbf{W}^o_k) &= p(\mathbf{w}^t_k|\mathbf{w}^o_k, \mathbf{W}^o_{k-1}) \\
&= \frac{p(\mathbf{w}^t_k, \mathbf{w}^o_k, \mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k, \mathbf{W}^o_{k-1})} \\
&= \frac{p(\mathbf{w}^o_k|\mathbf{w}^t_k, \mathbf{W}^o_{k-1})p(\mathbf{w}^t_k, \mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k, \mathbf{W}^o_{k-1})} \\
&= \frac{p(\mathbf{w}^o_k|\mathbf{w}^t_k, \mathbf{W}^o_{k-1})p(\mathbf{w}^t_k|\mathbf{W}^o_{k-1})p(\mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k|\mathbf{W}^o_{k-1})p(\mathbf{W}^o_{k-1})} \\
&= \frac{p(\mathbf{w}^o_k|\mathbf{w}^t_k, \mathbf{W}^o_{k-1})p(\mathbf{w}^t_k|\mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k|\mathbf{W}^o_{k-1})}.
\end{aligned}$$

This relates the transition probability of interest with pdf's that can be calculated more promptly.

Whiteness of the observation sequence allows us to write

$$p(\mathbf{w}_k^o|\mathbf{w}_k^t, \mathbf{W}_{k-1}^o) = p(\mathbf{w}_k^o|\mathbf{w}_k^t)$$

and therefore,

$$p(\mathbf{w}_k^t|\mathbf{W}_k^o) = \frac{p(\mathbf{w}_k^o|\mathbf{w}_k^t)p(\mathbf{w}_k^t|\mathbf{W}_{k-1}^o)}{p(\mathbf{w}_k^o|\mathbf{W}_{k-1}^o)}$$

It remains for us to determine each one of the transition probability densities in this expression.

Assumption: all pdf's (processes) are Gaussian and the observation process is linear, that is, $\mathbf{w}_k^o = \mathbf{H}_k\mathbf{w}_k^t + \mathbf{b}_k^o$, with $\mathbf{b}_k^o \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$.

In this case, an immediate relationship between the variables above and those from the example of estimating a constant vector can be drawn:

- $\mathbf{z} \rightarrow \mathbf{w}_k^o$
- $\mathbf{w} \rightarrow \mathbf{w}_k^t$
- $p_{\mathbf{z}|\mathbf{w}}(\mathbf{z}|\mathbf{w}) \rightarrow p(\mathbf{w}_k^o|\mathbf{w}_k^t)$
- $p_{\mathbf{w}}(\mathbf{w}) \rightarrow p(\mathbf{w}_k^t|\mathbf{W}_{k-1}^o)$
- $p_{\mathbf{z}}(\mathbf{z}) \rightarrow p(\mathbf{w}_k^o|\mathbf{W}_{k-1}^o)$

Consequently we have

$$p(\mathbf{w}_k^o|\mathbf{w}_k^t) = \frac{1}{(2\pi)^{m_k/2}|\mathbf{R}_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^t)^T\mathbf{R}_k^{-1}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^t)\right]$$

where we noticed that

$$\mathcal{E}\{\mathbf{w}_k^o|\mathbf{w}_k^t\} = \mathcal{E}\{(\mathbf{H}_k\mathbf{w}_k^t + \mathbf{b}_k^o)|\mathbf{w}_k^t\} = \mathbf{H}_k\mathbf{w}_k^t$$

and

$$\begin{aligned} cov\{\mathbf{w}_k^o, \mathbf{w}_k^o|\mathbf{w}_k^t\} &\equiv \mathcal{E}\{[\mathbf{w}_k^o - \mathcal{E}\{\mathbf{w}_k^o|\mathbf{w}_k^t\}][\mathbf{w}_k^o - \mathcal{E}\{\mathbf{w}_k^o|\mathbf{w}_k^t\}]^T|\mathbf{w}_k^t\} \\ &= \mathbf{R}_k \end{aligned}$$

Analogously, we have

$$p(\mathbf{w}_k^o|\mathbf{W}_{k-1}^o) = \frac{1}{(2\pi)^{m_k/2}|\mathbf{\Gamma}_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_{k|k-1}^f)^T\mathbf{\Gamma}_k^{-1}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_{k|k-1}^f)\right]$$

where we define $\mathbf{w}_{k|k-1}^f$ and the $m_k \times m_k$ matrix $\mathbf{\Gamma}_k$ as

$$\mathbf{w}_{k|k-1}^f \equiv \mathcal{E}\{\mathbf{w}_k^t|\mathbf{W}_{k-1}^o\}, \quad \mathbf{\Gamma}_k \equiv \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k$$

with the $n \times n$ matrix $\mathbf{P}_k^f$ defined as

$$\mathbf{P}_{k|k-1}^f \equiv \mathcal{E}\{[\mathbf{w}_k^t - \mathbf{w}_k^f][\mathbf{w}_k^t - \mathbf{w}_k^f]^T|\mathbf{W}_{k-1}^o\}$$

To fully determine the *a posteriori* conditional pdf $p(\mathbf{w}_k^t|\mathbf{W}_k^o)$, it remains to find the *a priori* conditional pdf $p(\mathbf{w}_k^t|\mathbf{W}_{k-1}^o)$. Since we assumed all pdf's to be Gaussian, the from the definitions of $\mathbf{w}_k^f$ and $\mathbf{P}_k^f$ above we have $p(\mathbf{w}_k^t|\mathbf{W}_{k-1}^o) \sim \mathcal{N}(\mathbf{w}_{k|k-1}^f, \mathbf{P}_{k|k-1})$, that is,

$$\begin{aligned} p(\mathbf{w}_k^t|\mathbf{W}_{k-1}^o) &= \frac{1}{(2\pi)^{n/2}|\mathbf{P}_k^f|^{1/2}} \\ &\quad \exp\left[-\frac{1}{2}(\mathbf{w}_k^t - \mathbf{w}_{k|k-1}^f)^T(\mathbf{P}_{k|k-1}^f)^{-1}(\mathbf{w}_k^t - \mathbf{w}_{k|k-1}^f)\right] \end{aligned}$$

and the conditional pdf of interest can be written as

$$p(\mathbf{w}_k^t|\mathbf{W}_k^o) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}_{k|k}^a|^{1/2}} \exp\left(-\frac{1}{2}J\right)$$

where

$$J = (\mathbf{w}_{k|k}^a - \mathbf{w}_k^t)^T(\mathbf{P}_{k|k}^a)^{-1}(\mathbf{w}_{k|k}^a - \mathbf{w}_k^t)$$

is the cost function, with $\mathbf{w}_{k|k}^a$ minimizing it.

We can now identify the quantities $\hat{\mathbf{w}}_{\mathrm{MV}}$ and $\mathbf{P}_{\tilde{\mathbf{w}}}$ of the problem of estimating a constant vector with $\mathbf{w}_k^a$ and $\mathbf{P}_k^a$, respectively. Consequently, it follows from this correspondence that

$$\begin{aligned} \mathbf{w}_{k|k}^a &= \mathbf{w}_{k|k-1}^f + \mathbf{P}_{k|k-1}^f\mathbf{H}_k^T\mathbf{\Gamma}_k^{-1}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_{k|k-1}^f) \\ (\mathbf{P}_{k|k}^a)^{-1} &= (\mathbf{P}_{k|k-1}^f)^{-1} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k \end{aligned}$$

Remarks

- The estimate $\mathbf{w}^a_{k|k}$ maximizing the *a posteriori* pdf is the MAP estimate.

- Moreover, since the resulting *a posteriori* pdf is Gaussian, this estimate is also the conditional mean, that is,

$$\mathbf{w}^a_{k|k} \equiv \mathcal{E}\{\mathbf{w}^t_k|\mathbf{W}^o_k\},$$

  and therefore it is the MV estimate which is what the Kalman filter obtains.

- Similar results can be obtained by minimizing the cost function

$$J_{3dVar}(\delta\mathbf{w}_k) \equiv \delta\mathbf{w}_k^T(\mathbf{P}^f_{k|k-1})^{-1}\delta\mathbf{w}_k + (\mathbf{v}_k - \mathbf{H}_k\delta\mathbf{w}_k)^T\mathbf{R}_k^{-1}(\mathbf{v}_k - \mathbf{H}_k\delta\mathbf{w}_k)$$

  where $\delta\mathbf{w}_k \equiv \mathbf{w}^t_k - \mathbf{w}^f_{k|k-1}$, and $\mathbf{v}_k \equiv \mathbf{w}^o_k - \mathbf{H}_k\mathbf{w}^f_{k|k-1}$. In the meteorological literature $J_{3dVar}(\delta\mathbf{w}_k)$ is referred to as the incremental three-dimensional variational (3dvar) analysis cost function.

- Since in practice we have only rough estimates of the observations and forecast error covariance matrices $\mathbf{R}_k$ and $\mathbf{P}^f_{k|k-1}$, the minimization problem above solves none other than a LS problem, given some prior information.

Remarks (cont.)

- So far we have made no assumptions about the process $\mathbf{w}_k^t$ other than its conditional pdf $p(\mathbf{w}_k^t|\mathbf{W}_{k-1}^o)$ being Gaussian. However, if we want to be able to calculate an estimate of the state one time ahead, that is at t_{k+1}, using the knowledge gather up to time t_k we must consider the pdf

$$\begin{aligned} p(\mathbf{w}_{k+1}^t, \mathbf{w}_k^t|\mathbf{W}_k^o) &= p(\mathbf{w}_{k+1}^t|\mathbf{w}_k^t, \mathbf{W}_k^o)p(\mathbf{w}_k^t|\mathbf{W}_k^o) \\ &= p(\mathbf{w}_{k+1}^t|\mathbf{w}_k^t)p(\mathbf{w}_k^t|\mathbf{W}_k^o) \end{aligned}$$

  which refers to the yet unspecified transition pdf $p(\mathbf{w}_{k+1}^t|\mathbf{w}_k^t)$ and therefore we must know more about the process $\mathbf{w}_k^t$.

- When the process $\mathbf{w}_k^t$ is linear the calculations are simple. That is, the system

$$\mathbf{w}_{k+1}^t = \mathbf{M}_{k+1,k}\mathbf{w}_k^t + \mathbf{b}_{k+1}^t$$

  with $\mathbf{b}_{k+1}^t \sim \mathcal{N}(0, \mathbf{Q}_{k+1})$ results in a Gaussian transition pdf (for an initial Gaussian pdf $p(\mathbf{w}_0^t)$):

$$p(\mathbf{w}_{k+1}^t|\mathbf{w}_k^t) \sim \mathcal{N}(\mathbf{M}_{k+1,k}\mathbf{w}_k^t, \mathbf{Q}_{k+1})\,.$$

- For linear dynamical process above it follows that

$$\begin{aligned} \mathbf{w}_{k+1}^f &= = \mathbf{M}_{k+1,k}\mathcal{E}\{\mathbf{w}_{k+1}^t|\mathbf{W}_k^o\} + \mathcal{E}\{\mathbf{b}_{k+1}^t|\mathbf{W}_k^o\} \\ &= \mathbf{M}_{k+1,k}\mathbf{w}_{k|k}^a \\ \mathbf{P}_{k+1|k}^f &= cov\{\mathbf{w}_{k+1}^t, \mathbf{w}_{k+1}^t|\mathbf{W}_k^o\} \\ &= \mathbf{M}_{k+1,k}\mathbf{P}_{k|k}^a\mathbf{M}_{k+1,k}^T + \mathbf{Q}_{k+1} \end{aligned}$$

The Extended Kalman Filter (EKF)

$$\begin{aligned}
\mathbf{w}^f_{k|k-1} &= \mathbf{f}(\mathbf{w}^a_{k-1|k-1}) \\
\mathbf{P}^f_{k|k-1} &= \mathbf{F}_{k-1|k-1}\mathbf{P}^a_{k-1|k-1}\mathbf{F}^T_{k-1|k-1} + \mathbf{Q}_k \\
\mathbf{K}_{k|k} &= \mathbf{P}^f_{k|k-1}\mathbf{H}^T_{k|k-1}\mathbf{\Gamma}^{-1}_{k|k-1} \\
\mathbf{P}^a_{k|k} &= \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_{k|k-1}\right)\mathbf{P}^f_{k|k-1} \\
\mathbf{w}^a_{k|k} &= \mathbf{w}^f_{k|k-1} + \mathbf{K}_{k|k}\left(\mathbf{w}^o_k - \mathbf{h}(\mathbf{w}^f_{k|k-1})\right)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{F}_{k-1|k-1} &\equiv \mathbf{F}(\mathbf{w}^a_{k-1|k-1}) = \left.\frac{\partial \mathbf{f}(\mathbf{w})}{\partial \mathbf{w}^T}\right|_{\mathbf{w}=\mathbf{w}^a_{k-1|k-1}} \\
\mathbf{H}_{k|k-1} &\equiv \mathbf{H}(\mathbf{w}^f_{k|k-1}) = \left.\frac{\partial \mathbf{h}(\mathbf{w})}{\partial \mathbf{w}^T}\right|_{\mathbf{w}=\mathbf{w}^f_{k|k-1}}
\end{aligned}$$

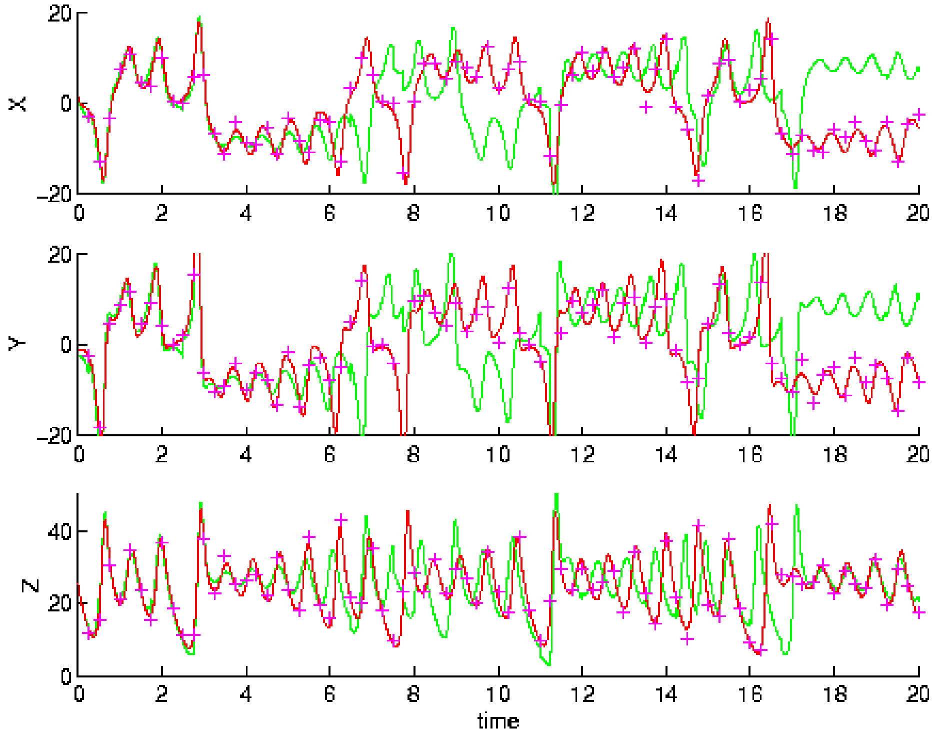
and

$$\mathbf{\Gamma}_{k|k-1} \equiv \mathbf{H}_{k|k-1}\mathbf{P}^f_{k|k-1}\mathbf{H}^T_{k|k-1} + \mathbf{R}_k$$

Illustration 1(cont.): Data Assimilation for Chaotic Dynamics

Then, what does data assimilation do?

σ(obs) = 2

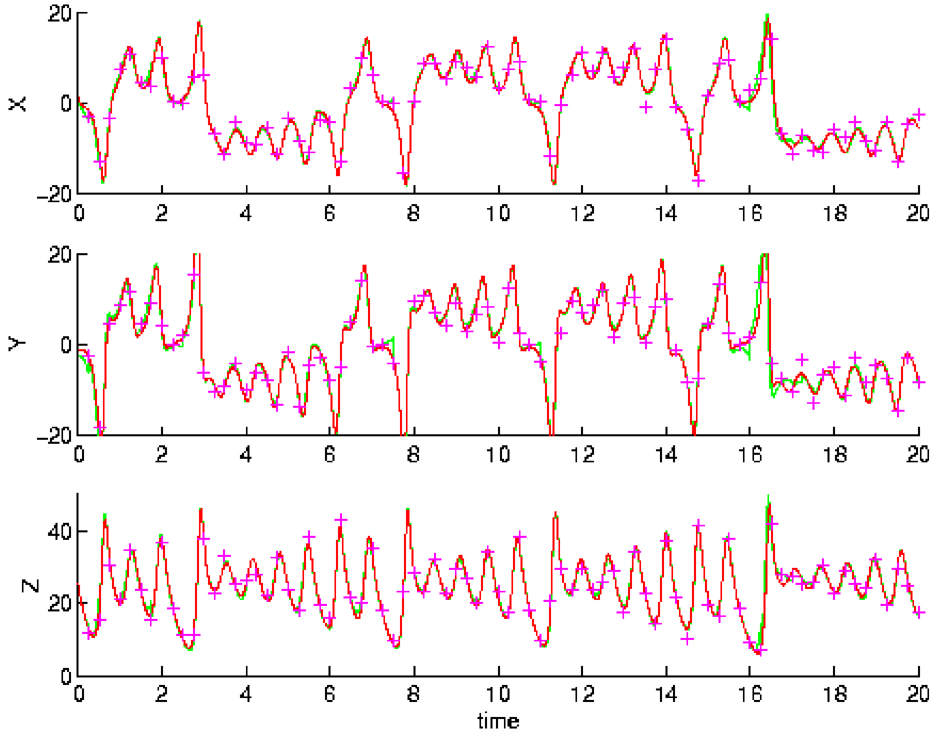


<u>Answer</u>: It improves our ability to estimate the true state and make relatively reasonable short- to medium-range predictions. However, depending on the data assimilation scheme, the estimate may diverge after a while. The red line represents the true state while the green line represents the estimate (assimilation), the crosses are the observations; the data assimilation scheme is the extended Kalman filter (EKF).

Illustration 1(cont.): Data Assimilation for Chaotic Dynamics

What if the data assimilation scheme is improved?

σ(obs) = 2

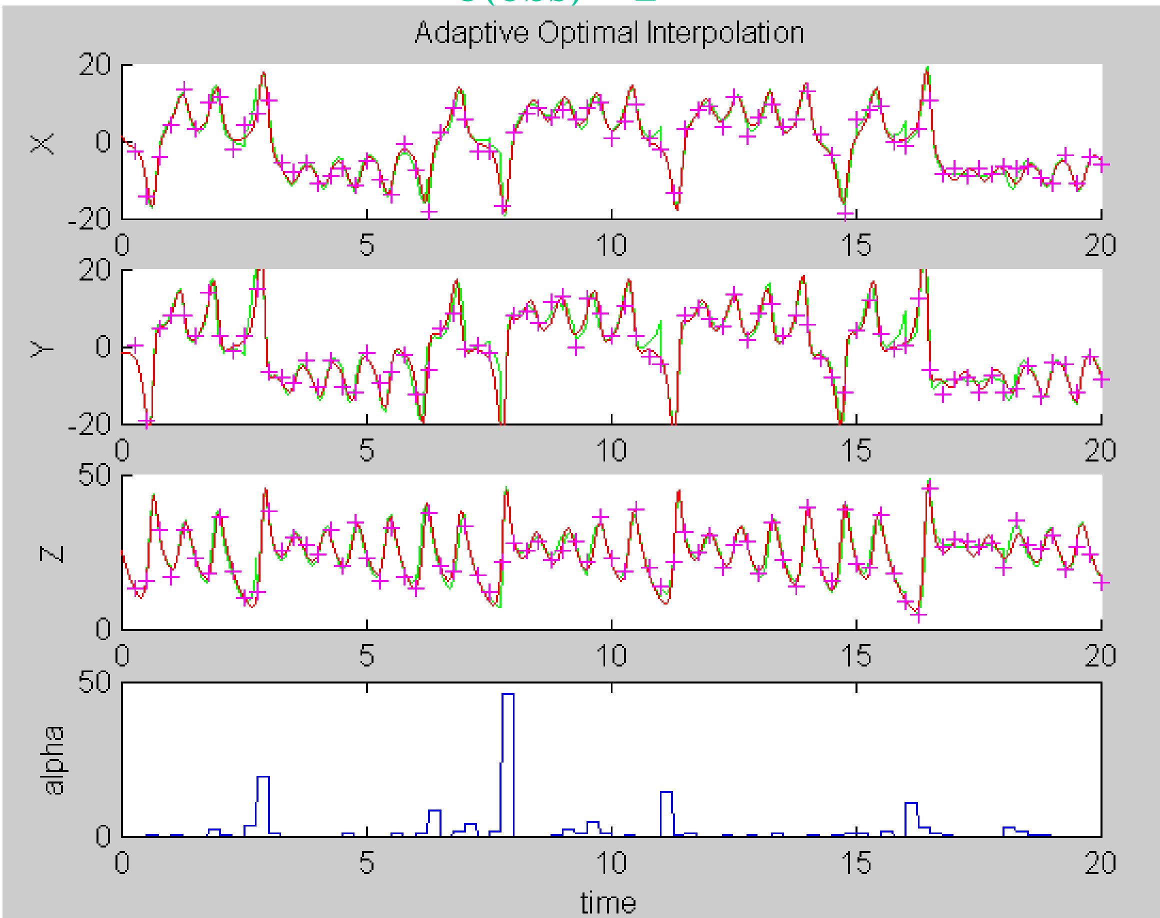


Answer: We get great results! Here we estimate the error due to linearization in the EKF via a Monte Carlo procedure proposed by Miller et al. (1994). We calculate the model error covariance off-line and then add that to the on-line EKF assimilation procedure. This is a rather good solution to prevent the EKF divergence due to misrepresentation of nonlinearities. However, this is a bit impractical for large data assimilation systems.

Illustration 1(cont.): Data Assimilation for Chaotic Dynamics

How does a simplified assimilation scheme perform?

σ(obs) = 2

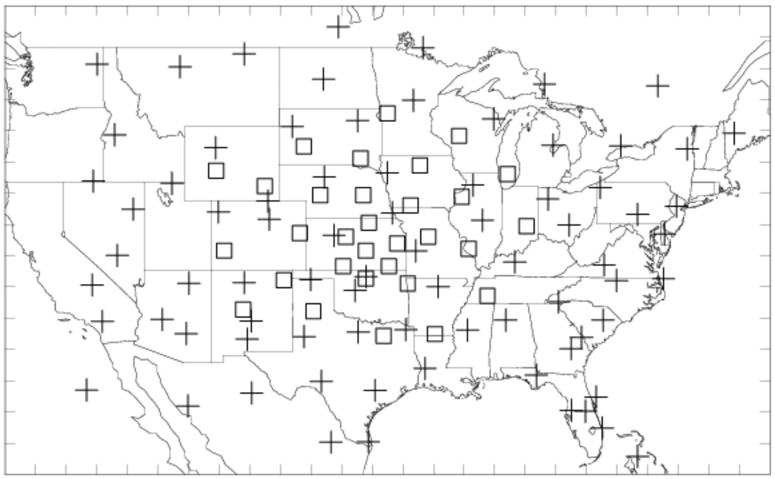


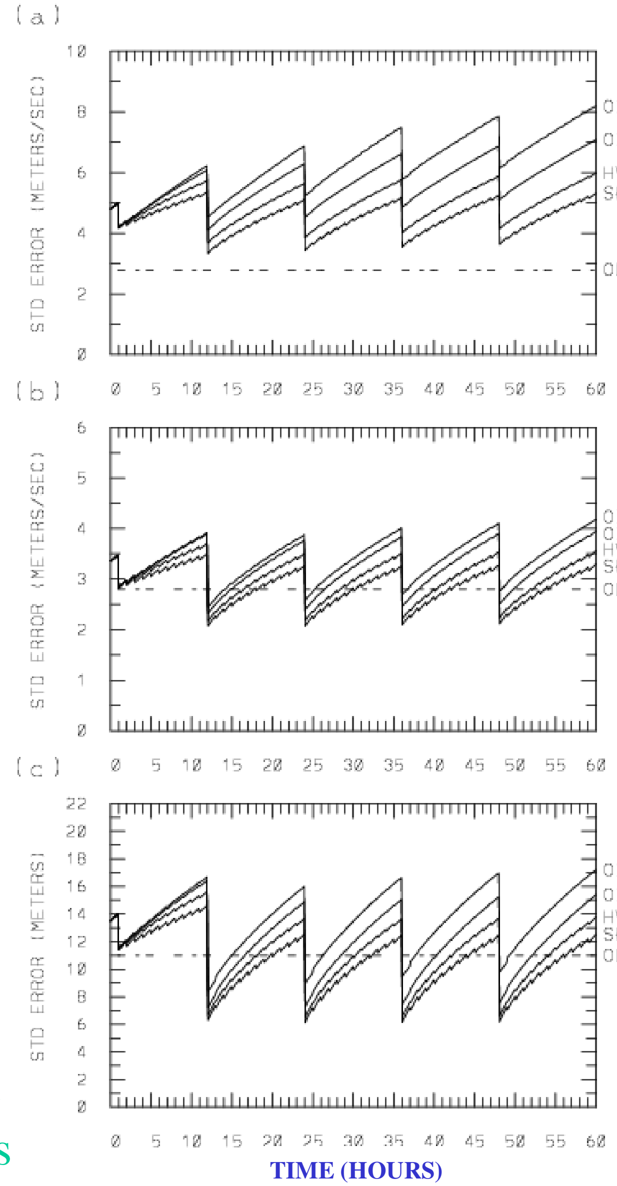
Answer: **Quite well! The assimilation scheme here is an adaptive optimal interpolation. In this case, the propagated error covariance (the costly part of the EKF) is replaced by a constant forecast error covariance matrix scaled by a single parameter that gets to be adaptively estimated on the basis of the observation-minus-forecast residuals (see Dee 1995). The time series of this estimated parameter is displayed in the lower panel above.**

Illustration 2: Linear shallow-water stable (nearly normal) dynamics

State Variables: U, V, H

Observations: 6-hourly radiosondes (plus signs)
hourly wind profilers (squares)





U-error

V-error

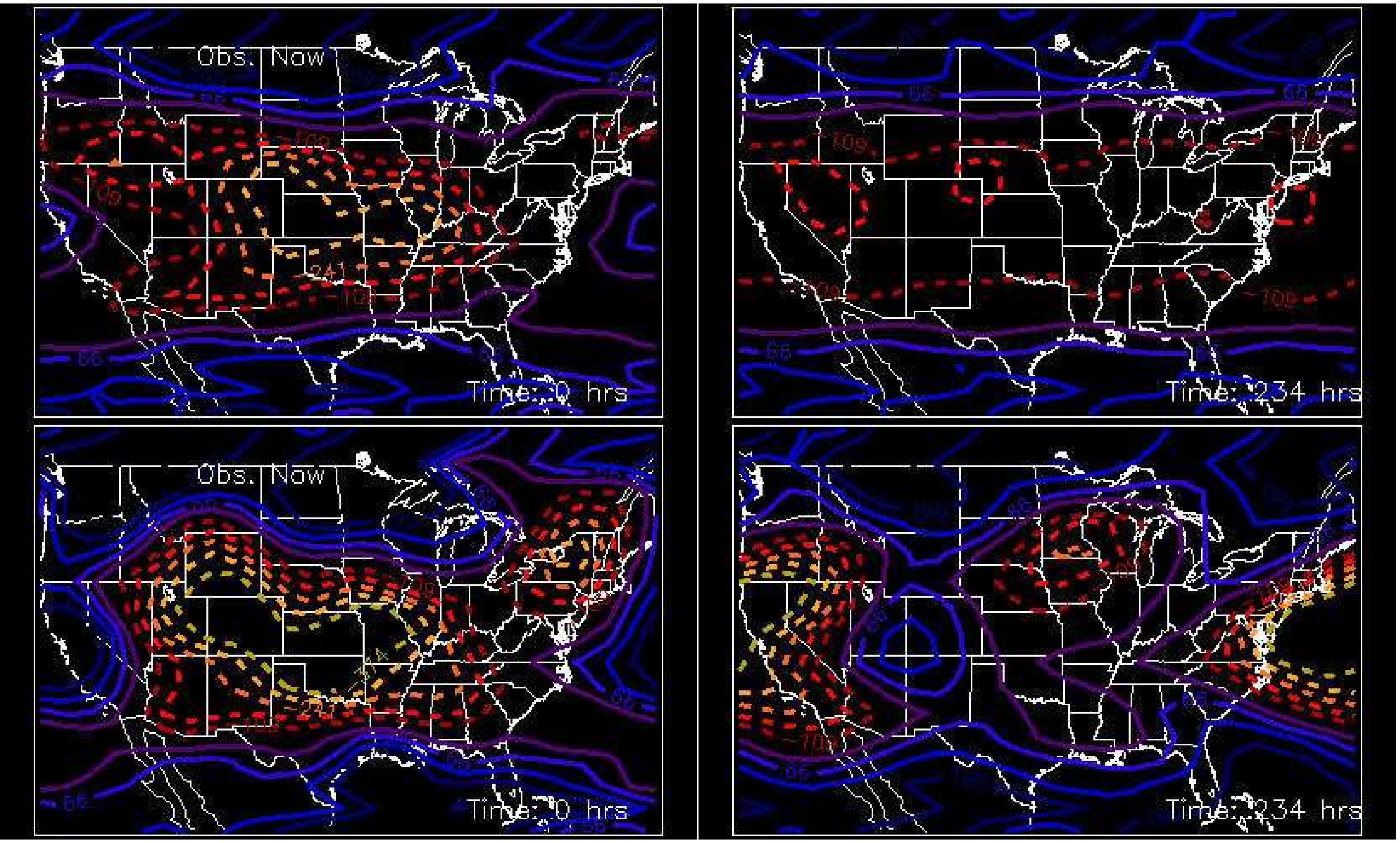
H-error

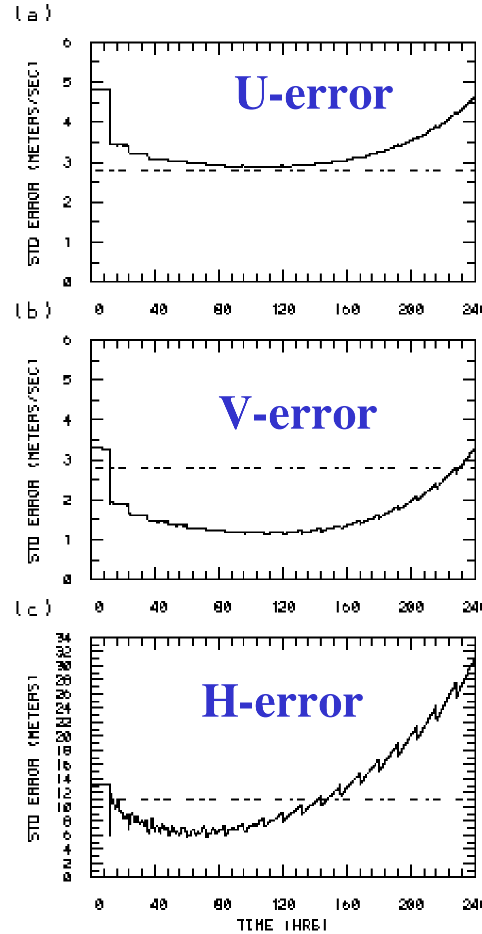
The results of a variety of simplifications to the KF have been studied for this model. The Fig. on the right shows the expected RMS error in estimating each of the system's variables for each scheme. Slow improvements are seem from a simple optimal interpolation scheme (OID) with no account for forecast error propagation to a scheme that propagates only the geopotential forecast error covariance (SKF). The SKF performs nearly as well as the KF (not shown in the figures).

Illustration 3: The Destructive Effect of Gravity Waves in Data Assimilation

T=0: True H(top) & Estimate
Contour interval is the same

T=234hrs: True H(top) & Estimate
Controur interval is still the same!!



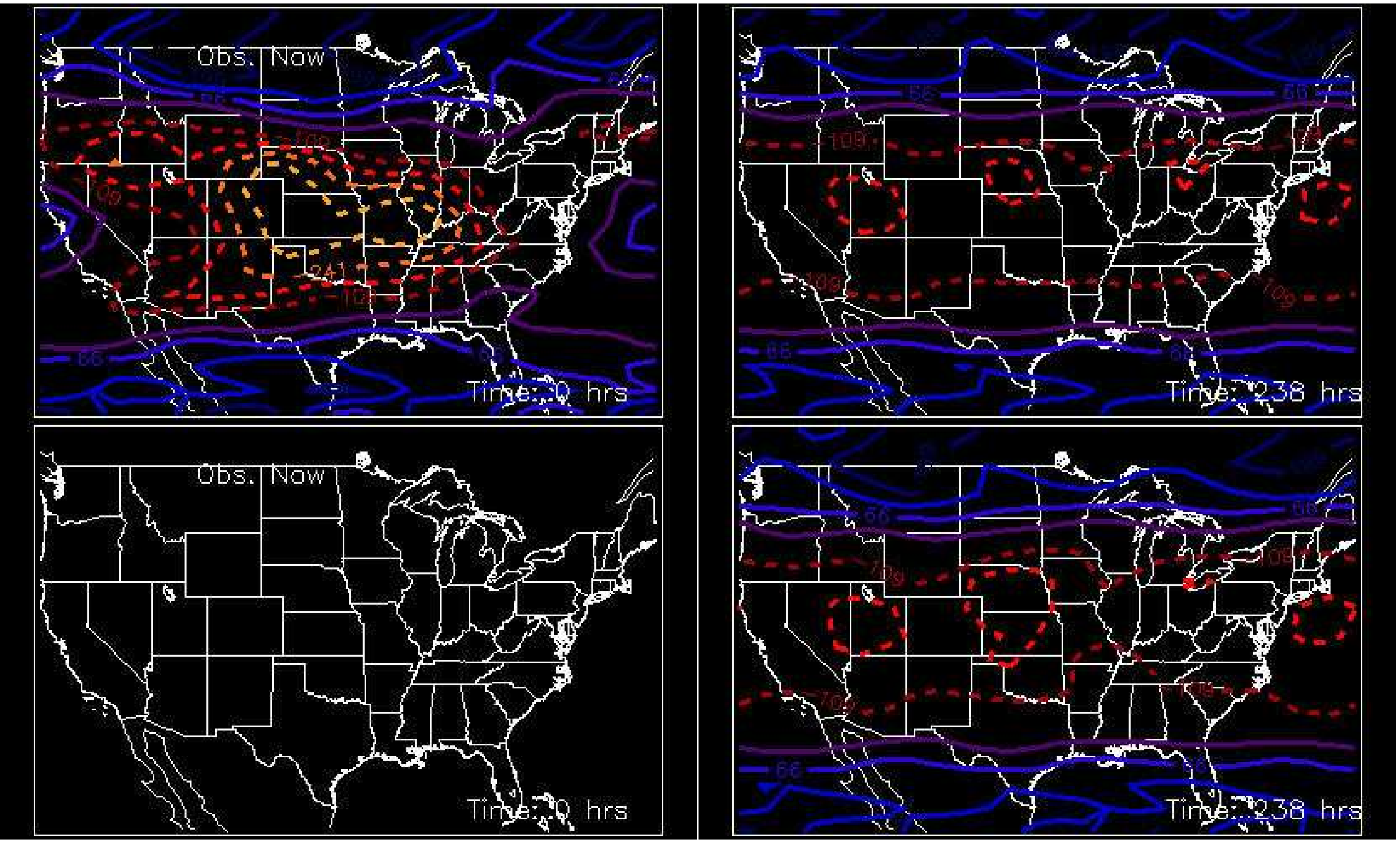


We now apply an OI scheme (an approximation to the Kalman filter) to assimilate observations into an unstable shallow-water dynamics. The OI scheme has no initialization and its error covariances are not the best. The lack of balance in the OI estimates excites gravity waves, which prevent the state from being properly estimated. Eventually the intrinsic dynamical instabilities evolve differently for the true state and the OI estimate and their difference grows exponentially (see ERMS error plots on the right for the three fields). The left most picture shows the true (top) and OI estimate mass field at t=0 – differences here are sole due to lack of precise knowledge of initial condition; the middle plot shows these quantities at t=234hrs when the true state and the OI estimate are completely different. WATCH THE MOVIE.

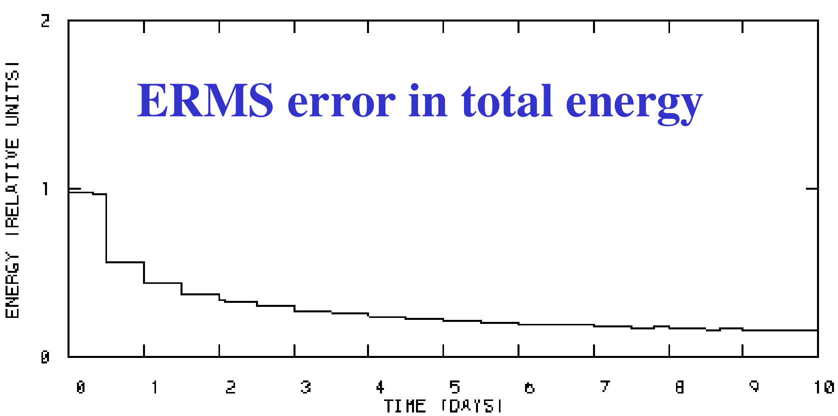
Illustration 4: Wave Generation

T=0: True H(top) & Estimate
Contour interval is the same

T=234hrs: True H(top) & Estimate
Controur interval is still the same!!



We now apply the KF to a case of wave-generation. This is done by assuming the initially (t=0) our estimate for the state of the system is zero (bottom-left) panel), which is clearly far from what the true state looks like (top-left). After some time the observations bring the estimate to within good agreement with the true state through the assimilation procedure. WATCH THE MOVIE.



The figure on the left shows the expected root-mean-square error in total energy as a function of time. This indicates that as time passes the difference between the estimate gets dramatically reduced – even though our original estimate is so far fetched.

5. *The Probabilistic Approach to Smoothing*

Smoothing is the problem of determining the state of a system given all the data available before, during, and after the time of the desired estimate. In this respect, the smoothing problem refers to the following conditional pdf

$$p(\mathbf{W}_k^t|\mathbf{W}_N^o) = p(\mathbf{w}_1^t, \cdots, \mathbf{w}_{k-1}^t, \mathbf{w}_k^t|\mathbf{w}_1^o, \cdots, \mathbf{w}_{N-1}^o, \mathbf{w}_N^o)$$

where $N \geq k$.

Remarks

- In general, an estimate obtained by maximizing the pdf

$$p(\mathbf{w}_k^t|\mathbf{w}_1^o, \cdots, \mathbf{w}_{k-1}^o, \mathbf{w}_k^o)$$

will be distinct from one maximizing $p(\mathbf{W}_k^t|\mathbf{W}_N^o)$ above.

- However, in the linear, gaussian, white noise case, with $N = k$, maximization of either one of the pdf's above amounts to the same solution, at the final time t_k.

- When the error (noise) statistics are Gaussian, the pdf $p(\mathbf{W}_k^t|\mathbf{W}_N^o)$ is also Gaussian and its maximization amounts to minimization of the following quadratic cost function:

$$J_N = \sum_{i=0}^{N} ||\mathbf{w}_i^o - \mathbf{H}_i\mathbf{w}_{i-1}||_{\mathbf{R}_i^{-1}} + \sum_{i=0}^{N} ||\mathbf{w}_i - \mathbf{M}_{i,i-1}\mathbf{w}_{i-1}||_{\mathbf{Q}_i^{-1}}$$

with respect to the entire trajectory $\mathbf{w}_0, \mathbf{w}_1, ..., \mathbf{w}_N$, and subjected to the ICs: $\mathbf{M}_{0,-1}\mathbf{w}_{-1} = \bar{\mathbf{w}}_0$ and $\mathbf{Q}_{-1} = \bar{\mathbf{P}}_0$.

▷ Minimization of the cost function J_N solves the fixed-interval smoother. In some sense, this is the problem that 4Dvar attempts to solve.

There are different ways of solving the smoother problem sequentially. The fixed-lag Kalman smoother is a particulaly attractive formulation.

To examplify consider the case of seeking an improved state estimate at t_{k-1} given observations up to time t_k. The relevant pdf is

$$p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_k) = \frac{p(\mathbf{w}^o_{k-1}|\mathbf{w}^t_{k-1})p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_{k-1})}{p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_k)} \,.$$

When all pdf's are Gaussian we can show that the maximum probability is obtained by minimizing

$$\begin{aligned} J_{lag=1} &= (\mathbf{w}^o_k - \mathbf{H}_k\mathbf{M}_{k,k-1}\mathbf{w}^t_{k-1})^T\tilde{\mathbf{R}}_k^{-1}(\mathbf{w}^o_k - \mathbf{H}_k\mathbf{M}_{k,k-1}\mathbf{w}^t_{k-1}) \\ &+ (\mathbf{w}^a_{k-1|k-1} - \mathbf{w}^t_{k-1})^T(\mathbf{P}^a_{k-1|k-1})^{-1}(\mathbf{w}^a_{k-1|k-1} - \mathbf{w}^t_{k-1}) \end{aligned}$$

where $\tilde{\mathbf{R}}_k \equiv (\mathbf{H}_k\mathbf{Q}_{k-1}\mathbf{H}_k^T + \mathbf{R}_k)$, and the optimal solution is found to be

$$\begin{aligned} \mathbf{w}^a_{k-1|k} &= \mathcal{E}\{\mathbf{w}^t_{k-1}|\mathbf{W}^o_k\} \\ &= \mathbf{w}^a_{k-1|k-1} + \mathbf{P}^a_{k-1|k-1}\mathbf{M}^T_{k,k-1}\mathbf{H}_k\mathbf{\Gamma}_k^{-1}(\mathbf{w}^o_k - \mathbf{H}_k\mathbf{w}^f_{k|k-1}) \,. \end{aligned}$$

Lag-ℓ Retrospective Data Assimilation Algorithm

for $k = 1, 2, \ldots$

$\mathbf{w}^f_{k|k-1} = f(\mathbf{w}^a_{k-1|k-1})$

$\mathbf{v}_{k|k-1} = \mathbf{w}^o_k - h(\mathbf{w}^f_{k|k-1})$

$\mathbf{\Gamma}_k = \mathbf{H}_k \mathbf{P}^f \mathbf{H}^T_k + \mathbf{R}_k$

Solve PSAS: $\mathbf{\Gamma}_k \mathbf{x} = \mathbf{v}_{k|k-1}$

$\mathbf{y}^0 = \mathbf{H}^T_k \mathbf{x}$

Analysis Increment: $\delta \mathbf{w}^a_{k|k} = \mathbf{P}^f \mathbf{y}^0$

for $\ell = 1, 2, \ldots, \min(k, L)$

Adjoint Integration: $\mathbf{y}^\ell = \mathbf{A}^T_{k-\ell|k-\ell} \mathbf{y}^{\ell-1}$

Solve PSAS: $\mathbf{\Gamma}_{k-\ell} \mathbf{x} = \mathbf{H}_{k-\ell} \mathbf{P}^f \mathbf{y}^\ell$

$\mathbf{y}^\ell := \mathbf{y}^\ell - \mathbf{H}^T_{k-\ell} \mathbf{x}$

Retro–Analysis Increment: $\delta \mathbf{w}^a_{k-\ell|k} = \mathbf{P}^f \mathbf{y}^\ell$

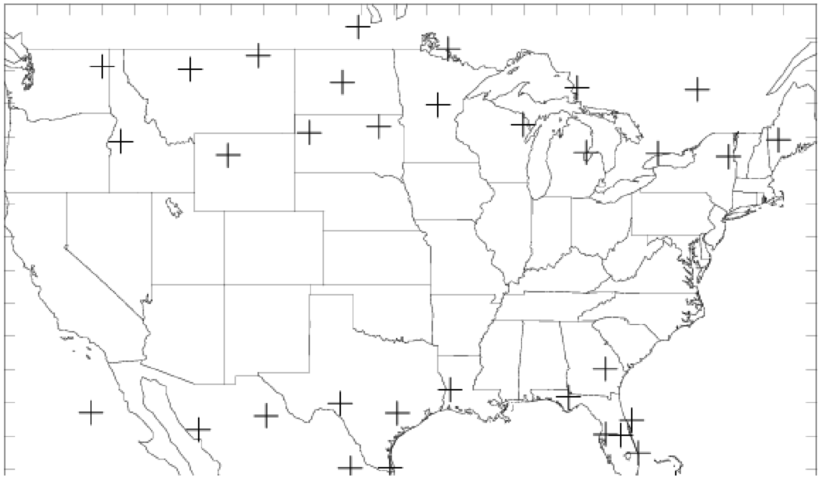
endfor

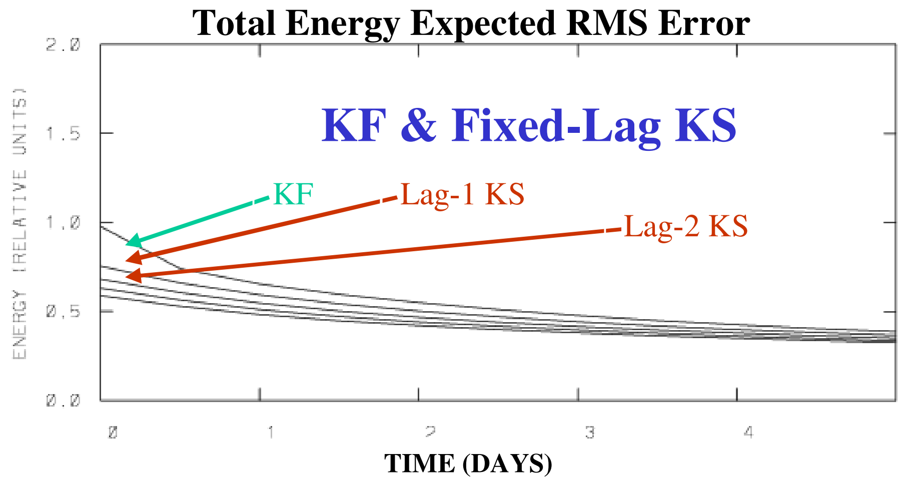
endfor

Note: matrix A on this page is the same as M on the previous page.

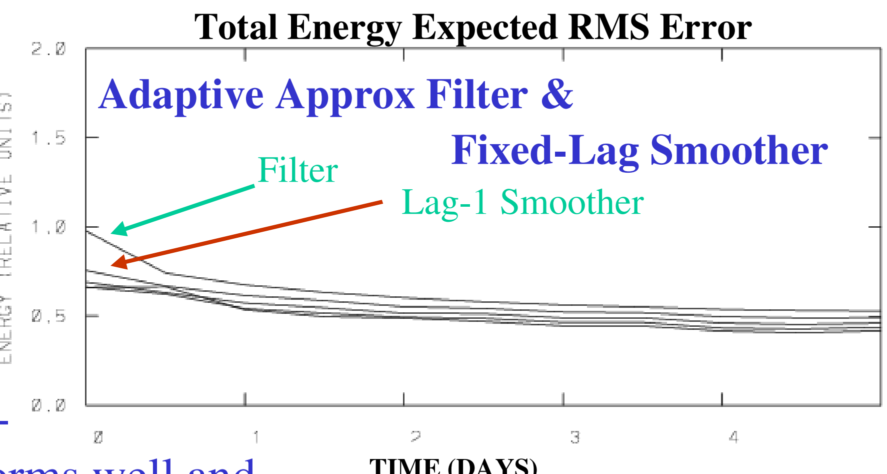
Illustration 5: Approximations to Fixed-Lag Smoothing

Now the SW has been turned in to an unstable dynamics by having it linearized around a zonal jet. To make the true state harder to estimate no observations are made available over the central (most important) region in the domain where the wind jet is strongest.



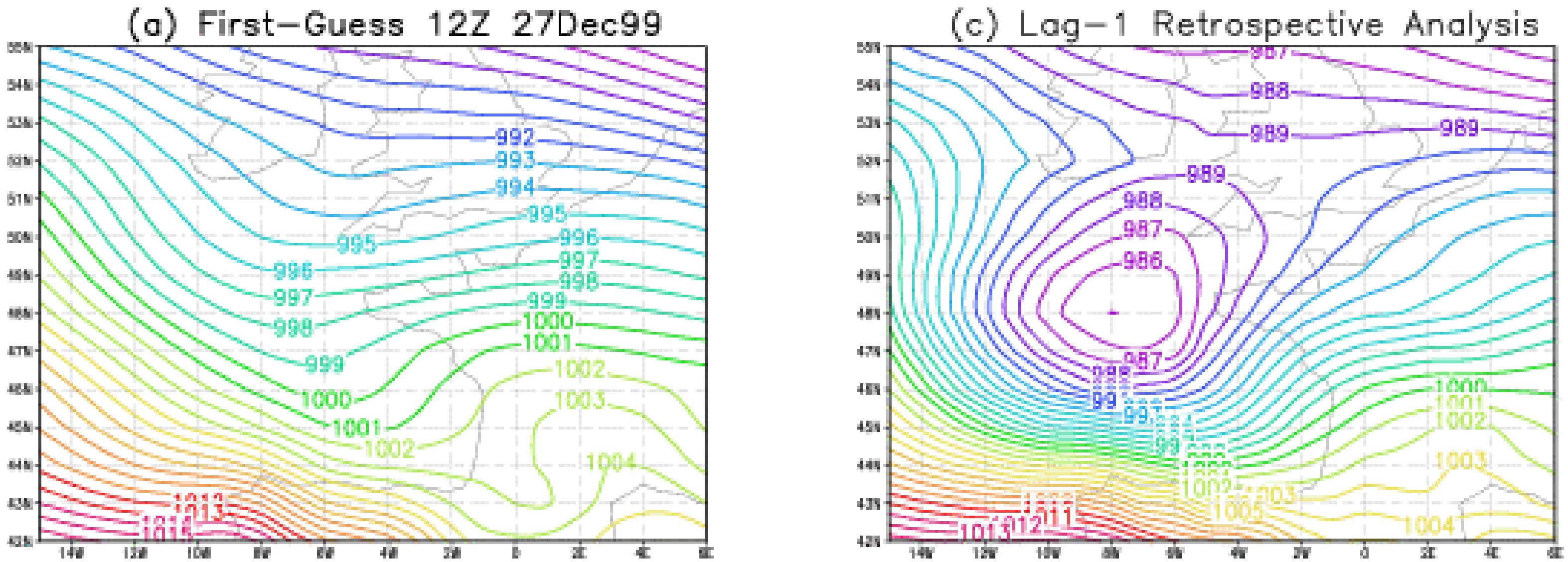


The to- right panel shows that even for unstable dynamics the KF is stable and so is the FLKS. Moreover, the FLKS shows improvement over KF for each lag (results for up to lag-4 are displayed).



The bottom-right panel shows that an adaptive approximate filter is also stable (robust). A FL-Smoother build on the basis of such filter performs well and Shows improvement over filter.

Illustration 6: Fixed-Lag Smoother: Improving the analysis of synoptic events



Case study: French storm of 27Dec1999 – implementation of a fixed-lag smoother in the DAO data assimilation system. Panels (c) and (d) show improvement in low of sea level pressure indicative of the storm.

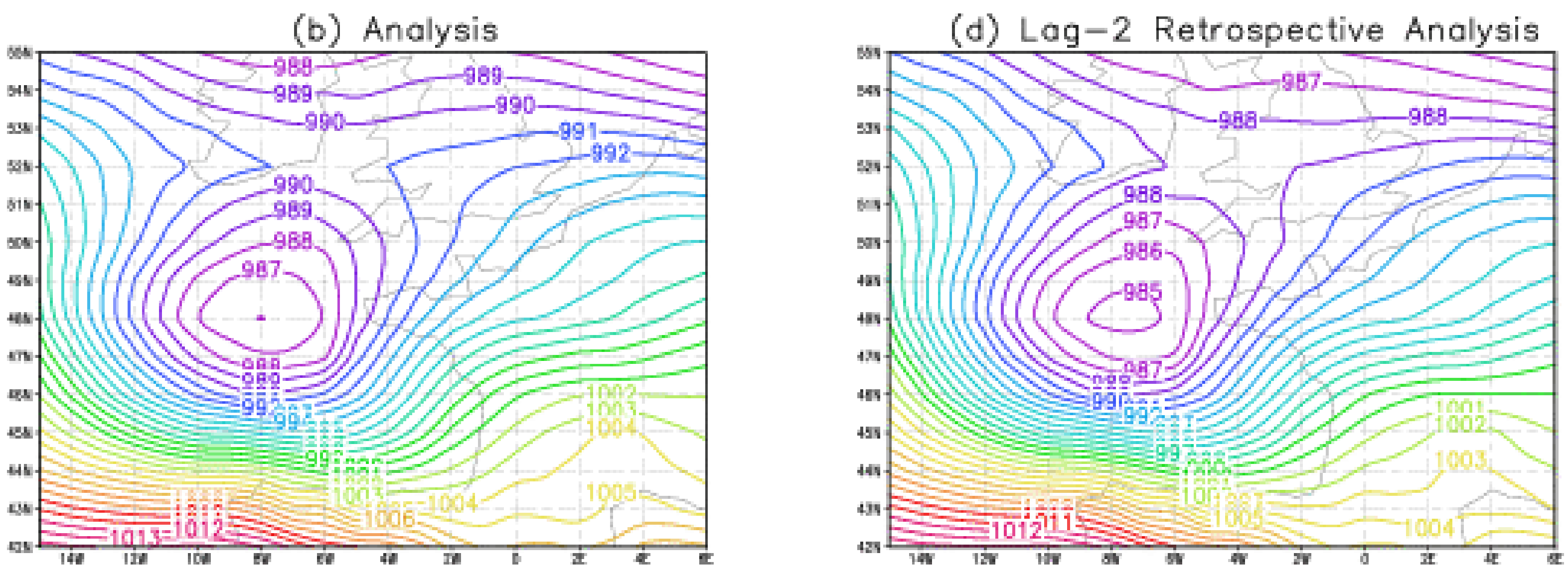
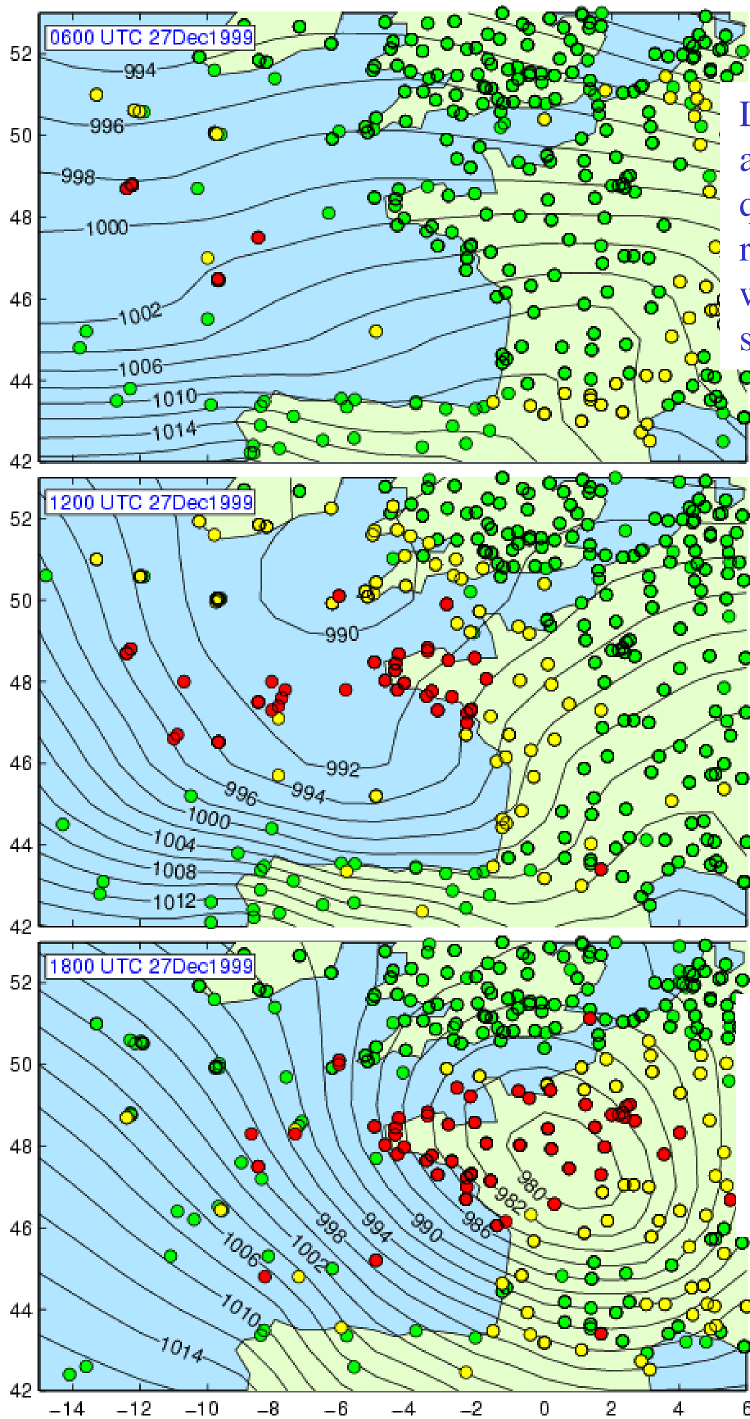
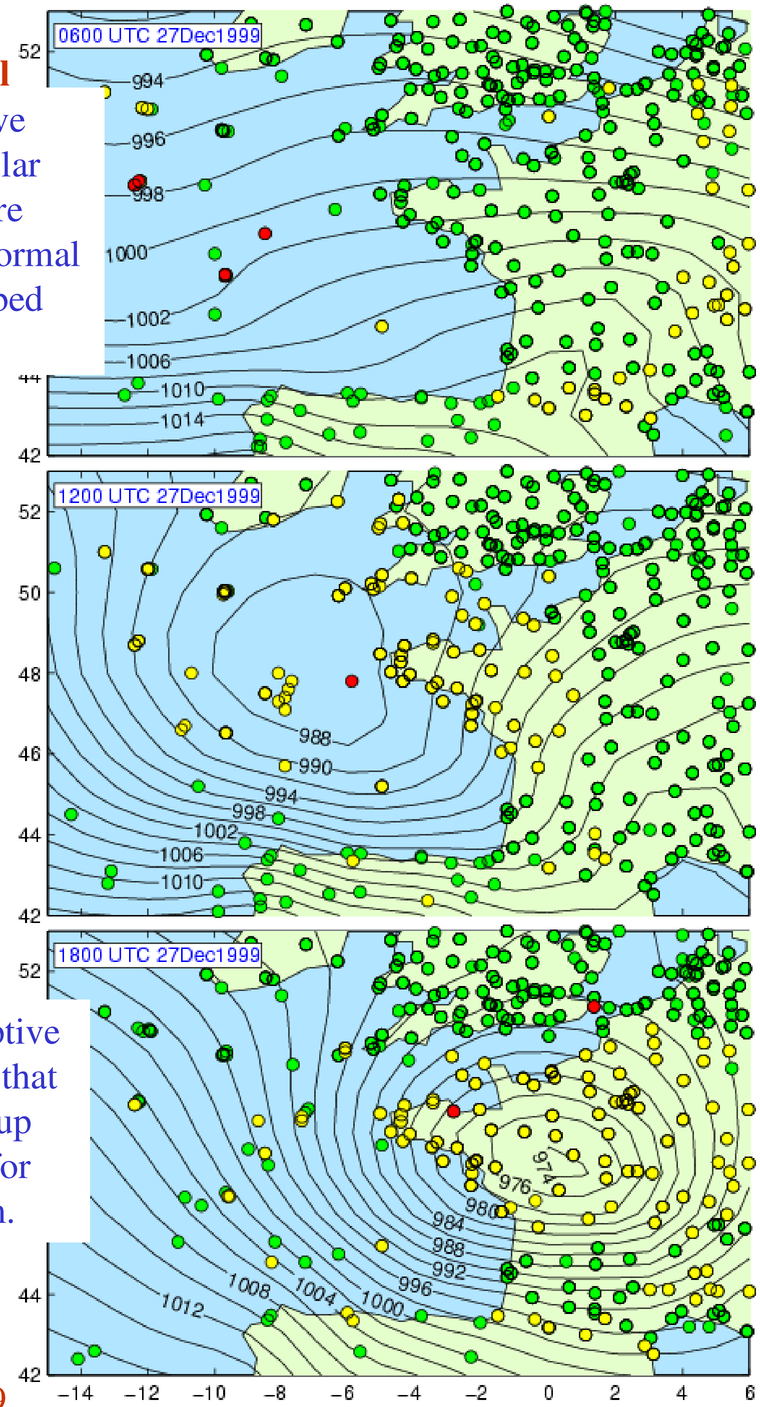


Illustration 7: Adaptive Quality Control



Left panels show consecutive analyses obtained with regular quality control: many obs are rejected since the seem abnormal when in contrast w/ prescribed statistics.



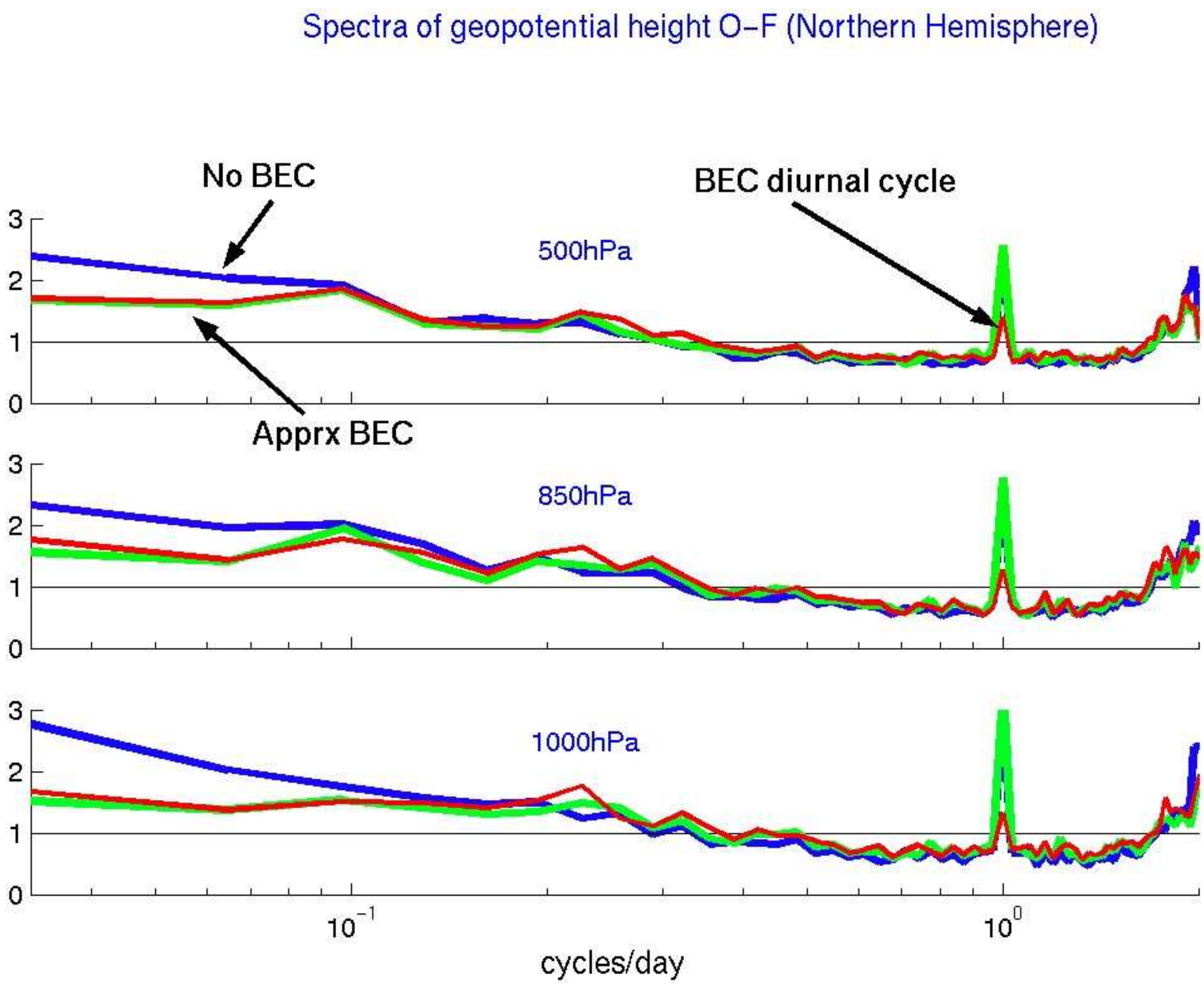
When the QC uses an adaptive buddy-check observations that seem suspect at first wide up being taken and allowing for better analysis of the storm.

Case study: French storm of 27Dec1999

Illustration 8: Diurnal Cycle Bias Estimation and Correction

$$\mathbf{b}_k^f = \mathbf{c}_k^0 + \mathbf{c}_k^1 \cos \omega t + \mathbf{d}_k^1 \sin \omega t$$

$$\text{for } \omega = 2\pi/24$$



Estimation theory can be used to estimate and correct for model biases.

The example here shows the spectra of the obs-minus-forecast residuals when forecast biases are present (**blue**) and when they are estimated and removed (**green**).If the filtering scheme behind these residuals is optimal the residuals will show a flat spectra around one indicative of them being white; notice how the tail of the green spectrum is lower than that of the blue spectrum.

Providing a model for the bias “evolution” allows removal of, say, biases in the diurnal cycle. Notice reduction of peak at 1cycle/day (**red**) when the model written above is used.

Closing Thought

Most of the methods to solve inverse problems are either Least-Squares or bear a close relationship to Least-Squares.

So, my advise to someone just starting in this field is to learn well about Least-Squares; what it means; and how it relates to methods like Kalman filtering/smoothing, and 3d/4d variational procedures.

Iterative methods to solve matrix-vector problems are often employed when calculating the LS-like solution an estimation problem. So, learn well conjugate-gradient, Newton methods, etc.

General remark: It helps to fix notation. One attempt to set standard notation came out of the Data Assimilation for Atmosphere and Oceans in Japan in 1995.
See Ide et al., in the Special Issue of the J. Meteorol. Soc. Japan of 1997.

Short (very biased) Reference List

Anderson, B.D.O., & J.B. Moore, 1979: *Optimal Filtering*. Prentice-Hall, 357 pp.

Cohn, S.E., 1997: An introduction to estimation theory. M. Ghil, K. Ide, A. Bennett, P. Courtier, M. Kimoto, N. Nagata, & N. Sato (Eds.): *Data Assimilation in Meteorology and Oceanography: Theory and Practice*, Universal Academic Press, 147-178.

Courtier, P., 1997: Dual formulation of four-dimensional variational assimilation. Q. J. Roy. Meteor. Soc., Part B, 123, 2449-2461.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.

Dee, D.P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. Mon. Wea. Rev., **123**, 1128-1145.

Dee, D.P., & R. Todling, 2000: Data assimilation in the presence of forecast bias: the GEOS moisture analysis. *Mon. Wea. Rev*., 128, 3268-3282.

Ghil, M., & P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Advances Geophys*., Vol. 33, Academic Press, 141-266.

Jazwinski, A.H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.

Menard, R., and R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. Tellus, **48A**, 221-237.

Tarantola, A., 1994: *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier, 613 pp.

Todling, R., 1999: *Class Notes on Estimation Theory and Atmospheric Data Assimilation*. NASA/DAO Office Note 99-01, 187 pp.

Todling, R., & S.E. Cohn, 1994: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. Mon. Wea. Rev., **122**, 2530-2557.

Todling, R., S.E. Cohn, & N.S. Sivakumaran, 1998: Suboptimal schemes for retrospective data assimilation based on the fixed-lag Kalman smoother. Mon. Wea. Rev., **126**, 2274-2286.

Wunsch, C., 1996: *The Ocean Circulation Inverse Problem*. Cambridge University Press, 442 pp.